PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:	WO 92/07080
C12N 15/85, 15/87	A1	(43) International Publication Date:	30 April 1992 (30.04.92)

(21) International Application Number: PCT/US91/07690

(22) International Filing Date: 16 October 1991 (16.10.91)

(30) Priority data:
598,881 17 October 1990 (17.10.90) US

(71) Applicant: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM [US/US]; 201 West 7th Street, Austin, TX 78701 (US).

(72) Inventors: HEWITT, Roger, R. ; 2415 White Oak Drive, Houston, TX 77009 (US). SICILIANO, Michael, J. ; 12462 Barryknoll, Houston, TX 77024 (US). LEGERSKI, Randy, J. ; 2425 Underwood, #341, Houston, TX 77030 (US).

(74) Agent: PATTERSON, Melinda, L.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).

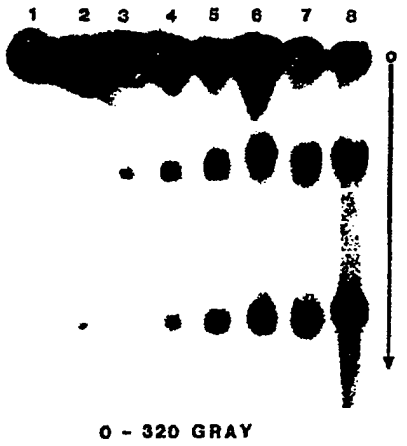
(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: A EUKARYOTIC EPISOMAL DNA CLONING AND EXPRESSION VECTOR

(57) Abstract

This invention relates to the development of recombinant eukaryotic cloning and expression vectors based on unique regulatory elements isolated from autonomously replicating, stable episomal units from human tumor cell lines. More specifically, the unique regulatory elements relate to origins of replication, as well as conferring extrachromosomal stability and maintenance. This cloning and expression vector will accommodate genes that exceed the cosmid limit (greater than 50 kb) and permit their maintenance as autonomously replicating extrachromosomal elements in mammalian cells. Inclusion of telomeres and centromeres would control the replication and segregation and therefore serve as an eventual vehicle for gene replacement therapy. This invention is therefore unique in providing for the expression and autonomous replication of large genes, maintained extrachromosomally, in a vector containing episomal regulatory elements.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU+	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE*	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

+ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

A EUKARYOTIC EPISOMAL DNA CLONING AND EXPRESSION VECTOR

This invention relates to the development of recombinant eukaryotic cloning and expression vectors based on unique regulatory elements isolated from autonomously replicating, stable episomal units isolated from human tumor cell lines. More specifically, the unique regulatory elements include origins of DNA replication, and DNA sequences that confer extrachromosomal stability and maintenance. These unique episomal regulatory elements permit large pieces of DNA to be expressed or cloned (greater than 50 kilobase pairs [kb] in size).

During the past decade, the underlying significance of recent advances in molecular biology has been the ability to clone and manipulate DNA from virtually any source by ligating restriction fragments into phage or plasmid vectors which are then replicated in *E. coli*.

Since then, a crucial technological gap has developed in what is commonly called "conventional recombinant DNA technology." This technological gap stems from two developments. The first is the discovery that many eukaryotic genes are encoded by enormous lengths of DNA. The second is an optimistic and enthusiastic goal of mapping and sequencing entire genomes, including the human genome.

Because of the large size of DNA in many genes from higher organisms, this size limitation and restriction can be stifling. For example, bithorax locus in *Dropsophila,* which plays an active role in the fly's segmentation pattern, encompasses approximately 320 kb (Karch, et al., Cell 43:81, 1985). Factor VIII gene in the human which encodes the blood-clotting factor deficient in hemophiliacs, spans at least 190 kb (Gitschier, et al., Nature (London), 312:326, 1984). The gene that is defective in Duchenne's muscular dystrophy is estimated to include more than a million base pairs (1000 kb). A striking feature of this gene is the protein-coding portion may be encoded by as little as 15 kb of DNA (Monaco, et al, Nature (London), 302:575, 1983). Thus, there is a strong need for technological advances which permit the cloning and expression of very large genes.

Also widening this technological gap is the increased interest in and enthusiasm for gene replacement therapy. Proposals to use genes to treat cancer and immune deficiencies have only recently been approved by the National Institutes of Health human gene therapy subcommittee and the Recombinant DNA Advisory Committee (Science, 249:974, August, 1990). These first studies focus on:

(1) delivering tumor necrosis factor (TNF) directly to a tumor site in much larger doses by packaging the gene for TNF inside special lymphocytes that have a natural affinity for tumors; and

(2) attempting actual gene replacement therapy in children with a rare, inherited and often lethal immune system disorder caused by adenosine deaminase deficiency. A normal healthy

recombinantly produced ADA gene will be introduced into the white blood cell of an ADA deficient child and the cells are then returned to the patient (Id. 975).

To narrow this gap, molecular biologists are attempting to clone large pieces of exogenous DNA into compatible hosts by means of artificial vectors. However, standard recombinant DNA techniques, that involve the construction of small plasmid vectors that can be transfected into host cells and clonally propagated, are limited in the amount of exogenous DNA that can be "squeezed" or inserted into these vectors. These size restrictions only permit about 50 kilobase pairs (kb) to be cloned into the vectors usually employed in cloning.

More limitations exist when the discussion turns to the bacterial expression of mammalian proteins. The current technology for expressing mammalian proteins in bacteria is hampered with problems relating to post translational modifications and functional bioactivity.

To date, cloning of large segments of exogenous DNA in the range of several hundred kilobase pairs has only been achieved by employing yeast. This was done by ligating exogenous DNA to vector sequences that allow their propagation as linear artificial chromosomes (Burke, et al, Science, 236:806, 1987). Although this technique is a significant step towards resolving this size restriction, cloning large segments of exogenous DNA into yeast is not without limitations. Questions and concerns about this technology pertain to (1) the stability of the recombinants, (2) whether clone banks are representative of the starting material, (3) whether the desired protein is consistently expressed in extrachromosomal vectors, and (4) whether normal human transcripts are properly

processed in yeast, as well as, whether proper expression and post translational modification of the recombinant protein occurs in yeast.

Therefore, with the yeast expression system and its limitations, there is still a very strong need to design and construct eukaryotic expression and cloning vectors possessing the capabilities of housing very large regions of DNA (greater than 50 kb) and of accurately processing and expressing of these large genes. With such a novel vector, large regions of DNA that span genes can then be cloned and whole proteins encoded by the genes can then be expressed.

One mechanism by which a cell can accumulate large amounts of specific protein or RNA is by amplification of the respective gene. This amplification may be located on either expanded chromosomal regions (homogenous staining regions) or on extrachromosomal autonomously replicating elements (called double minute, double minute chromosomes or episomes).

Episomes have unique features; the most notable are that episomes autonomously replicate and are stably maintained extrachromosomally. The characteristics of episomes permits the continuous production of the respective amplified gene and the gene products it encodes. For example, an episome produced in hamster cells has been characterized to contain amplified amounts of a transfected CAD (CAD is an acronym for the multifunctional protein containing carbamylphosphate synthetase, aspartate transcarbamylase, and dihydroorotase) gene at high frequency (Carrol, et al., Molecular and Cellular Biology, 7(5):1740, 1987). The amplified CAD gene is produced with each division of each cell.

Viral episomes have also been identified. It has been demonstrated that papilloma viral DNA replicates like a plasmid in mouse cells. Circular bovine papilloma virus (BPV) DNA can transform certain mouse cell lines to a malignant phenotype. In these transformed cell lines, the BPV DNA remains circular and extrachromosomal at about 30 - 100 copies per cell. This "plasmid" is being stably maintained in higher eukaryotes. Desired genes may be inserted into the BPV DNA and be maintained in the plasmid-like state and high levels of mRNA and protein corresponding to the desired gene can be produced. It has also been shown that Epstein-Barr virus vectors contain sequences that provide extrachromosomal stability of episomal DNA as well as origins of replication. This viral vector has been used to identify human DNA sequences that permit autonomous replication in human cells (Krysan, et al., Molecular and Cellular Biology, 9(3):1026, 1989). But, it can be appreciated that there are many limitations when working with a virally produced protein. For example, in terms of producing proteins that may ultimately be used to replace defective human genes, viral episomes probably are not feasible because of potential Food and Drug Administration regulations, etc. Also the viral episome eventually integrates into chromosomal sites which then interferes with continued amplification and causes the expression of its resident genes to be extinguished.

Thus, the limitations in terms of integration into chromosomal sites and of potential hazards pertaining to the use of viral based vectors for amplification and expression apply to all eukaryotic viral episomes.

It is the intent of this invention to describe a eukaryotic cloning and an expression vector which will accommodate genes that exceed the cosmid limit (greater

than 50 kb) and permit their accumulation and maintenance as autonomously replicating extrachromosomal elements in mammalian cells. This invention is therefore unique by providing autonomous replication and expression of large genes in a vector containing episomal regulatory elements.

This minimal cloning or expression vector will be further modified by the inclusion of regions of human chromosomes containing telomeres and centromeres. This would thus create a human artificial chromosome that would be subjected to the same control mechanisms (regarding regulation and chromosomal segregation) as normal chromosomes and therefore serve as a vehicle for gene replacement therapy. This modification of the extrachromosomal vector is therefore unique in that it will be a synthetic chromosome containing genes of choice, that will be expressed, and that will be maintained and regulated as if it were a normal chromosome.

This cloning or expression vector may take on several forms. For example, two principal forms for employment are: (1) employed via extrachromosomal/episomal, autonomous replication and segregation which could even be amplified, and (2) employed via a human artificial chromosome under normal chromosomal control mechanisms.

In general and overall scope, the present invention relates to the development of recombinant eukaryotic cloning and expression vectors based on unique regulatory elements isolated from autonomously replicating, stable episomal units isolated from human tumor cell lines. More particularly, these unique regulatory elements include origins of DNA replication, and DNA sequences that confer extrachromosomal stability and maintenance. These unique episomal regulatory elements will permit large pieces of

DNA to be expressed or cloned (greater than 50 kilobases pairs in size).

This invention discloses procedures for producing two different types of vectors. One is a cloning vector and the other one is an expression vector. For the purpose of this invention, the phrase "cloning vector" refers to a DNA vector designed to be used to clone a desired gene. The techniques that are involved in cloning vary from vector to vector and from system to system, however, these techniques in general are standard and known to those skilled in the art of recombinant DNA technology.

Also, for the purpose of this invention, the phrase "expression vector" refers to a DNA vector capable of replication in selected mammalian host cells and expressing a desired protein. This protein may then be recovered from the cells by employing techniques known to those skilled in the art.

This cloning vector should include one or more functional origins of DNA replication to permit stable, autonomous replication. The phrase "origin of replication" is defined as a region that indicates the origin of replication.

This cloning vector should include appropriate DNA sequences that confer extrachromosomal stability and maintenance. The sequences responsible for conferring extrachromosomal stability and persistence may be related to sequences responsible for nuclear matrix attachment sites, topoisomerase II reaction sites, and/or other regions required for appropriate interactions with the nuclear architecture. This extrachromosomal stability and maintenance permits the introduction of large exogenous genes.

This cloning vector should also include DNA selectable marker sequences that can be used to confer drug resistance to a transfected cell or DNA sequences that can correct a genetic mutation. This allows the cells that were transfected with the vector to be selected for. The DNA selectable marker segment confers upon a cell transfected with said vector, the ability to survive in the presence of a selected compound or selected group of compounds. The compound may be either G418 or hygromycin B. Also, other selectable marker segments will contain DNA encoding an enzyme capable of functionally replacing a mutated enzyme so as to render the transfected cell resistant to said selected compound or selected group of compounds. The enzyme may be selected from a group consisting of: thymidine kinase, xanthineguanine phosphoribosyl transferase, adenine phosphoribosyltransferase, adenosine deaminase and dihydrofolate reductase.

This cloning vector should also include a multi-use multiple cloning site to facilitate recovery for genetic modification and analysis and insertion for reintroduction into cells for replication and expression. Multiple cloning cassette sequence cartridges are commercially available from several different companies (Promega, New England Biolabs, etc). A typical cassette sequence would include restriction sites for 8 - 11 different enzymes (i.e. Eco RI, Sac 1, Sma 1, Ava I, Bam H1, Xba 1, Hinc II, Acc 1, Sal 1, Pst 1, Hind III, etc.) The availability of these cassette sequences are known to those skilled in the art.

This cloning vector should also include a DNA segment encoding bacterial components necessary for propagation of said vector in bacteria. Bacterial components that are essential for propagation of the cloning vector in

bacteria are known to those skilled in this art. For example, two bacterial components essential for bacterial propagation are a replicon that is responsible for initiation of replication and antibiotic resistant markers (i.e. ampicillin, tetracycline, etc.) that permits growth in specific antibiotics.

In addition to the above described five different components included in the unique cloning vector, a unique expression vector capable of expressing large pieces of DNA (40 - 400 kb) should also include, a promoter, a polyadenylation site and a splice site in spacial relation to allow efficient expression of a structural gene.

The choice of promoters to be included in this vector will depend on the mammalian host cell employed. It is advantageous to employ a compatible promoter with regard to the cells that the desired protein will be expressed in. The inventors prefer to employ promoters derived from the following genes (although other promoters would be satisfactory): cytomegalovirus, SV-40, Rous sarcoma virus, thymidine kinase, beta-actin, metallothionein, and the epidermal growth factor receptor gene isolated from a DiFi episome.

For the purpose of this invention, a polyadenylation site refers to the site at which a poly A tail (a stretch of 50 to 300 adenines) is added to the vector for efficient expression of a desired protein in a mammalian cell. Also, the phrase "splice site" refers to a bacterial processing site essential to remove introns incorporated into the bacterial plasmid. These components are essential for optimal expression of a desired protein.

A further embodiment of this invention is an artificial chromosome consisting of a DNA segment derived

from a non-viral episome, said segment containing an origin for DNA replication, a DNA segment derived from a non-viral episome, said segment containing a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector, a DNA segment containing a multiple cloning site, a DNA selectable marker segment conferring upon a cell transfected with said vector, the ability to survive in the presence of a selected compound or selected group of compounds, a DNA segment encoding bacterial components necessary for propagation of said vector in bacteria, a promoter, a polyadenylation site, a splice site, a DNA segment encoding a centromere and a DNA segment encoding a telomere.

Further in accordance for this invention is a substantially purified non-viral episome of human origin capable of stable extrachromosomal maintenance and of autonomous replication in a compatible mammalian cell line.

Further in accordance for this invention is a substantially purified episomal DNA segment containing an origin of replication. This invention further includes a substantially purified episomal DNA segment containing a DNA sequence, which confers upon a vector including said segment, the ability to be stably maintained extrachromosomally in a cell transfected with said vector. Another embodiment of this invention is a substantially purified episomal DNA segment containing both an origin of replication and a DNA sequence, which confers upon a vector including said segment, the ability to be stably maintained extrachromosomally in a cell transfected with said vector.

Another embodiment of this invention is a DNA segment containing an origin for DNA replication is from an episome isolated from DiFi colorectal cell line.

Another embodiment of this invention is a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector is from an episome isolated from DiFi colorectal cell line.

Another embodiment of this invention is a DNA segment containing an origin for DNA replication and a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector is from an episome isolated from DiFi colorectal cell line.

The various techniques which have been successfully applied to the cloning and expression of many genes in a variety of host systems, employing many different promoters and vectors, are known to those skilled in the art of recombinant DNA technology and could be applied to the embodiments described herein.

For the purpose of this invention, the phrase "operatively spaced with respect to a desired gene" is defined as the appropriate positional spacing required between the numerous cloning and expression vectors components described in this invention so as to allow each of the of components to achieve its desired function. These components are also directionally positioned 5′ to 3′. The appropriate spacing needed for efficient cloning or expression of a desired gene is determined for each individual vector.

In terms of transfecting eukaryotic cells with these unique cloning or expression vectors, the transfection techniques are standard and known to those skilled in the art of recombinant DNA technology. In terms of transfecting cells with the unique expression vector, this invention could also be applied for the production of stable cell lines which are, by definition, continuously producing the desired protein. The production of cell lines designed to continuously produce the desired protein has been described extensively in the literature, and is therefore known to those skilled in the art.

CHARACTERISTICS OF THE DEPOSITED CELL LINE

Cell line "DiFi" comprising cells obtained from the ascitic fluid of a colorectal tumor in a patient with Gardner's syndrome, is available from the ATCC, accession # CRL 10576. This cell line retains 50 copies or more of extrachromosomal episomes, each of which contains at least one complete copy of the epidermal growth factor receptor gene.

Fig. 1. In situ hybridization of DiFi cells with EGFR.

A portion of a metaphase from DiFi cells stained with Giemsa (A), fluorescence visualization of in situ hybridization using biotinylated EGFR as probe and counterstained with propidium iodide (B), and a black and white print of the fluorescence pattern of in situ hybridization (C).

Fig. 2. Electrophoretic mobilization of EGFR genes by gamma irradiation.

Autoradiogram of a Southern blot of a TAFE gel hybridized with 32P-labeled EGFR. Origin (o) is indicated at the top as is the direction of migration. Plug samples 1-8 were exposed to 0, 5, 10, 20, 40 80, 160, 320 Gray,

respectively. Hybridization membranes were exposed to film for 24 hrs.

Fig. 3. Effect of gamma irradiation on the electrophoretic mobilization EGFR in A431, DiFi, and HeLa cells.

Autoradiogram of a Southern blot of a TAFE gel hybridized with 32P-labeled EGFR. Origin and direction of migration is as in Fig. 2. A431, DiFi and HeLa cell DNA plugs were irradiated with A. OGy, B. 10 Gy, C. 40 Gy, D. 160 Gy. Autoradiographic exposure was extended to 72 hr in order to enhance sensitivity for detecting any fragments that might have migrated from the A431 plugs.

Fig. 4. CHEF analysis of EGFR in gamma irradiated DiFi.

Plugs containing DiFi DNA were exposed to 31.4 Gy prior to electrophoresis. The analysis of control (C) and irradiated (R) samples was performed in duplicate. Approximate sizes of the observed fragments, in kbs, are indicated to the right.

INTRODUCTION TO THE DISCLOSED INVENTION: AUTONOMOUSLY REPLICATING, STABLY MAINTAINED MICROCHROMOSOMAL UNITS FROM HUMAN TUMOR CELL LINES

In developing the invention, we elected to use stably maintained extrachromosomal units arising in some eukaryotic cell lines as starting material, because these units contain all the genetic regions required for autonomous replication and extrachromosomal expression. Those steps are described below.

In initial studies, the episomes are isolated from the origin in a substantially purified form and the minimal essential elements for episomal replication and transcription are localized and isolated. Those elements are then ligated into a selected DNA molecule, together

with additional DNA segments, including, for example, selectable markers, multiple cloning site or sites, segments necessary for propagation in bacteria and/or a promoter enhancer, splice site and polyadenylation site.

Replication of nuclear DNA in eukaryotes appears to be under precise and reproducible control, such that it is replicated only once in each S-phase, the DNA synthetic portion of each cell division cycle. In addition, each portion of the genome replicates at the same time in each S-phase, with expressing (transcribed) genes replicating early and non-expressing and/or structural DNA replicating late.

The genomes of prokaryotes, viruses, and yeast contain DNA sequences called origins, that serve as sites for initiating cycles of DNA replication. By analogy, such sites define replicating units, or replicons, in eukaryotic cells such as human cells.

An accepted working hypothesis is that the eukaryotic nucleus is organized into structural domains in which the nuclear matrix plays an essential role in organizing chromatin structure and in regulating function. Support for this hypothesis comes from studies demonstrating that DNA replication, DNA repair, transcription and post-transcriptional processing are associated with the nuclear matrix. Other studies have shown that DNA polymerase, RNA polymerase II, expressing and expressible genes, transcriptional enhancer sequences, topoisomerase II cleavage sites, topoisomerase II, and heterogeneous nuclear RNA (hnRNA) splicing complexes are highly enriched or specifically localized in the nuclear matrix.

The fact that regulatory DNA sequences and the nuclear proteins with which they interact have not been

identified is in part attributable to the unmanageable size of chromosomes and the complexity of the genetic elements they contain. However, stable cell lines are occasionally established in which regions of specific genes have been amplified (Stark, Cancer Surveys, 5:1-23, 1986) and occasionally are segregated into autonomously replicating components. These exist in the nucleus as episomes (200 kb - 800 kb molecules) and/or light microscope-visible double minute chromosomes (dmins, >1000 kb).

This invention exploits these cell lines by isolating and investigating the structure and replication control of their extrachromosomal elements in order to identify DNA sequences required to ensure their autonomy for stable maintenance, replication and gene expression. This minimal essential structure should then provide the core structure with which to assemble a cloning and expression vector for genes exceeding sizes accommodated by cosmid vectors.

Although the methodology described herein contains sufficient detail to enable one skilled in the art to practice the present invention, a commercially availbale technical manual entitled MOLECULAR CLONING (Maniatis, et. al., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York) may provide some additional details useful to assist practice of some aspects of this invention. Accordingly, this manual is incorporated herein by reference.

The following examples are designed to illustrate certain aspects of the present invention. However, they should not be construed as limiting the claims thereof.

EXAMPLE 1

EXTRACHROMOSOMAL AMPLIFICATION OF THE EPIDERMAL GROWTH FACTOR RECEPTOR GENE IN A HUMAN COLON CARCINOMA CELL LINE

This example describes the isolation and characterization of an autonomously replicating episomal unit derived from a human colorectal carcinoma cell, established from ascites from a patient with Gardner's syndrome, designated "Difi" (Bowman, et al., In: Hereditary Colorectal Cancer, J. Utsunomiya and H. Lynch (Eds.), Springer-Verlag, In Press, 1990). The invention is not limited to the "Difi" episome, however, for the basic procedures provided by the present disclosure should enable those of skill in the art to develop vectors from the episomes of other cells.

DiFi cells were (1) successfully established in tissue culture, (2) shown to contain amplified EGFR genes and mRNA, and (3) characterized cytologically to be near tetraploid with the presence of double minutes (dmin; Bowman et al. In Hereditary Colorectal Cancer, J. Utsunomiya and H. Lynch (eds), SpringVerlag, In Press, 1990).

I. CELL LINES EMPLOYED AND CELL CULTURE CONDITIONS

A431 (obtained from Gary Gallick, M. D. Anderson Cancer Center), HeLa and DiFi cells were maintained in Dulbecco's medium supplemented with 5% fetal and 5% newborn calf serum. SW480 cells, a colon tumor cell line (established by Leibovitz, 1976 and obtained from Mark Blick, M. D. Anderson Cancer Center) were grown and maintained in L-15 medium containing L-glutamine and supplemented with 10% fetal calf serum, insulin (5ug/ml) and glutathione (16ug/ml).

A. Characteristic of a human colorectal cancer cell line (DiFi)

"DiFi" colorectal carcinoma cell line represents one of the first cell lines to be established and characterized from a patient with Gardner syndrome.

Malignant ascitic fluid cells were isolated from a 46 year old female rectal cancer patient with Gardner syndrome and initiated to grow in culture. The cells have been maintained in culture for over three years. Hoechst stain analysis for mycoplasma was negative. Subcutaneous injection of DiFi cells into athymic mice demonstrated tumor production in 50% of the mice. The cells have a tetraploid karyotype, and possess an isozyme pattern characteristic of colorectal cancer cell lines.

II. LOCALIZATION OF EGFR DNA IN DiFi CELLS BY IN SITU HYBRIDIZATION

The following studies demonstrated the episomal location of the amplified EGFR gene.

Slides containing metaphase cells from either DiFi or SW480 cells were prepared and stored at room temperature. Prior to in situ hybridization with a biotinylated EGFR probe, the slides were stained (six minutes in 5% Giemsa prepared in phosphate buffer pH 6.8) and photographed. In situ hybridization involved treating the photographed slides with RNAse, DNA denaturation and dehydration solutions, overnight incubation in a hybridization mix containing a biotinylated EGFR probe, and tagging the regions of EGFR hybridization with fluorescein-avidin and biotinylated goat anti-avidin. This procedure resulted in a three layers of fluorescein-avidin, and visualization by fluorescence microscopy (Pinkel et al., Proc. Natl. Acad. Sci. USA, 83:2934, 1986).

The EGFR cDNA probe, HER-A64-3 (Ullrich et al., Nature 309:418-425, 1984), was labeled by nick translation with biotin-7-dATP according to the instructions provided by BRL. Hybridization mix (25 ul) containing 10% PEG 6000 and 5 ng of probe was used on each slide. Following in situ hybridization and fluorescence labeling procedures, slides were rinsed and counterstained in propidium iodide (2 ug/ml in H_20) for two minutes, rinsed with H_20, and carefully blotted dry. Two drops of antifade solution (Johnson and Aroujo, J. Immunol. Methods 43:349-350, 1981) were added to each slide before covering with a coverslip. Metaphase chromosomes were photographed under epi-UV-illumination on Kodak Ektachrome 160 film using the Zeiss filter 25 combination 48 77 09.

Giemsa-stained metaphase chromosomes from DiFi cells revealed a background of extrachromosomal particles at the limit of optical resolution (Fig. 1A). Occasionally, they were paired in the form of small dmins. To determine whether these structures contained copies of the EGFR gene, the biotinylated A64-3 cDNA EGFR probe was hybridized to these metaphase cells. SW480 cells served as a negative control because their dmins are amplified for MYC rather than EGFR (Untawale, Masters Thesis on File at the Graduate School of Biomedical Sciences, University of Texas Health Science Center, Houston, Texas, 1987; Untawale and Blick, Anticancer Res. 8:1-8, 1988). Thirty-five SW480 metaphase cells were examined for hybridization with biotinylated A64-3 cDNA EGFR probe. No hybridization was observed to any metaphase chromosome or extrachromosomal entity (data not shown). The same analysis was performed with DiFi metaphase spreads and thirty-three out of sixty-six demonstrated strong hybridization to extrachromosomal regions. No conclusions could be drawn from the remaining thirty-three metaphase cells due to weak hybridization or high background.

Figure 1 presents in situ hybridization of DiFi metaphase cells with EGFR probe. A portion of a metaphase spread from DiFi cells was stained with Giemsa (1A). Fluorescence visualization of in situ hybridization using biotinylated EGFR as a probe and counterstained with propidium iodide is shown in 1B, and a black and white print of the fluorescence pattern of in situ hybridization is shown in 1C.

In the Geimsa stained metaphase (1A) the chromosomes are intensely stained in contrast to the diffuse staining of extrachromosomal material in the background. The extrachromosomal background appears to be dmin, which vary in their size and visibility. Hybridization of the biotinylated EGFR probe (yellow fluorescence) was limited to extrachromosomal regions containing dmin, rather than chromosomal DNA (1B). In order to emphasize the extrachromosomal hybridization the photograph was printed in black and white (1C). In Figure 1C, the extrachromosomal labeling was visualized more clearly since the fluorescein fluorescence is more intense in dmin than isothe propidium fluorescence from the chromosomes.

Therefore, in situ hybridization of the biotinylated EGFR probe in the DiFi cell line demonstrated localized hybridization predominantly in extrachromosomal regions rather than to chromosomal DNA.

The in situ hybridization analysis presented in Figure 1B and 1C consistently demonstrated specific biotinylated EGFR localized in the extrachromosomal background. This specific localization is most likely associated with episomes many of which are too small in size and disorganized in structure to be visualized as dmins in standard cytogenetic spreads.

III. PREPARATION AND IRRADIATION OF DNA

After confirming that the EGFR amplification observed in the DiFi cells was mediated by a stable episomal fraction, we next sought to isolate that fraction from the cells using the procedures described below.

Cells were embedded, lysed and deproteinized in agarose blocks in order to minimize shear damage to the DNA (Smith et al., In Methods in Enzymology, M. Gottesman (Ed.), Academic Press, San Deigo, Vol 151, p. 461., 1987). Agarose blocks, with each sample containing approximately 3 ug of DNA, were cut to fit gel slots. Samples were suspended in 1 ml of TAFE buffer (10 mM Tris-acetate, pH 8.0; 0.5 mM EDTA) in 12 x 75 mm polystyrene culture tubes and exposed to ^{137}Cs gamma rays at a dose-rate of 45 Gray/min to linearize the DNA for pulse field electrophoresis (van der Blick et al., NAR 16:4841-4851, 1988; Beverly, NAR 16:925-939, 1988; Ruiz et al., Mol. Cell. Biol. 98:109-115, 1989). The inventors exposed agarose plugs containing unsheared DiFi cellular DNA to varying doses of gamma radiation prior to analysis by pulse-field gel electrophoresis. Appropriate levels of exposure were estimated based on an expected yield of 1.1 x 10^{-4} double-strand breaks/Gy/bp (calculated from Krisch et al., Rad. Res. 101:356-372, 1985).

IV. PULSED-FIELD GEL ELECTROPHORESIS WAS EMPLOYED TO SIZE DNA

Following irradiation, the samples were loaded into 1% agarose gels and subjected to transverse alternating field electrophoresis (TAFE) using TAFE buffer in a GeneLine system (Beckman Instruments). Agarose plugs containing yeast chromosomes or concatemers of lambda phage DNA were included on gels as size standards. Initial current was held constant at 170 ma for 30 min, reorienting the direction of the electrical field every 4

sec, followed by a constant current of 150 mA for 18 hr with a field reorientation interval of 60 sec.

Some experiments employed the clamped homogeneous electrical field (CHEF) protocol for pulsed-field gel electrophoresis (Chu et al., Science 232:65-68, 1986). Here, electrophoresis was performed in 0.5x TBE buffer (45 mM boric acid, 45 mM Tris and 2 mM EDTA, pH 8.3) at a constant current of 70 volts reoriented every 15 min for a total of 3 days.

V. SOUTHERN TRANSFER AND HYBRIDIZATION

Upon completion of electrophoresis, staining (0.5 ug/ml ethidium bromide), and photography, gels were irradiated for 5 min with 254 nm UVL (Gelman Instrument Co., Model 51438). This was followed by gentle shaking in 0.25 M HCl for 5 min for depurination, rinsing in deionized water, soaking in 0.4 M NaOH for 1 hr for hydrolysis of depurinated bases, rinsing in deionized water, and soaking in 0.2 M NaOH, 0.6 M NaCl for 1 hr for denaturation. The DNA was transferred to a Zetabind nylon membrane (AMF Cuno, Inc.) in the denaturing solution for 15-20 hrs. The filter was then treated with two 15 min washes in a neutralizing solution (0.5 M Tris-HCl, pH 7.5; 1.5 M NaCl) and dried in a vacuum oven at 80°C for 1 hr. Labeling of probe, hybridization to filters and autoradiography for visualization of fragments were performed as previously described (Amasino, Anal. Biochem. 152:304-307, 1986; Liu et al., Science 246:813-815, 1989).

Figure 2, an autoradiogram of a Southern blot of a TAFE gel probed with 32P-labeled EGFR, demonstrates electrophoretic mobilization of EGFR genes by gamma irradiation. The origin (o) as well as the direction of migration is indicated at the top of the figure. Plug samples 1-8 were exposed to 0, 5, 10, 20, 40 80, 160, 320

Gy, respectively. Hybridization membranes were exposed to film for 24 hrs.

Southern analysis of a gel hybridized with an EGFR probe demonstrated the dose dependent migration of two different sized fragments containing EGFR sequences (Fig. 2). The pattern of migration of total DNA was observed by staining gels with ethidium bromide (data not shown). Dose-dependent increases were observed in the amount of random sized DNA fragments migrating between the sample well and the front of each lane. Increased amounts of DNA also accumulated in the zone representing molecules of 2500 kb or larger under the electrophoresis conditions employed. The EGFR-containing fragments migrated at a position consistent with approximately 650 kb and 1300 kb representing faster and slower migrating forms, respectively. The origin is indicated by "O."

Figure 3, an autoradiogram of a Southern blot of a TAFE gel probed with 32P-labeled EGFR, demonstrates the effect gamma irradiation has on the electrophoretic patterns of migration of EGFR sequences in A431, DiFi, and HeLa cells. The origin and direction of migration are as in Fig. 2. DNA plugs from A431, DiFi and HeLa cells were irradiated with increasing amounts of radiation: Lane (A): OGy; Lane (B): 10 Gy; Lane (C): 40 Gy; Lane (D): 160 Gy. The autoradiographic exposure was extended to 72 hr in order to enhance sensitivity for detecting any fragments that might have migrated form the A431 plugs.

Dose dependent increases were observed in the amounts of randomly broken DNA fragments migrating from sample wells into each lane. As is observed, EGFR amplification is much higher in DiFi DNA and A431 DNA when compared to HeLa DNA. More importantly, sample plug irradiation did

not release discrete sizes of HeLa and A431 EGFR sequences were (confirmed by exposing autoradiograms for 7 days, data not shown). However, mobilization of both the 650 kb band and 1300 kb band DiFi EGFR fragments were readily detected. To summarize, EGFR sequences in both HeLa and A431 DNA appear to be chromosomally localized. In contrast, EGFR sequences in DiFi DNA appear to be episomally (extrachromosomally) localized and may be substantially purified by the procedure described here.

Figure 4 presents CHEF analysis of EGFR from gamma irradiated DiFi DNA. Plugs containing DiFi DNA were exposed to 31.4 Gy prior to electrophoresis. The analysis of control (c) and irradiated (R) samples was performed in duplicate. Approximate sizes of the observed fragments, in kbs, are indicated to the right. Irradiating DiFi plugs and conducting CHEF electrophoresis under conditions that resolve larger DNA fragments revealed the presence of a weakly hybridizing band of approximately 2,000 kb, in addition to the 650 kb and 1300 kb fragments (Fig. 4). In unirradiated control lanes (C) a small portion of EGFR-containing molecules were observed to have migrated into the gels. This observation was previously attributed to degradation of cellular DNA during the preparation of agarose plugs (van der Blick, et al., NAR 16:4841-4851, 1988).

VI. SUMMARY

In situ hybridization, using a biotinylated cDNA probe for the epidermal growth factor receptor (EGFR) gene, demonstrated that amplified EGFR in colon tumor cell lines, DiFi, is localized to many small double minute chromosomes of varying size and visibility. Analysis of the electrophoretic mobility of gammairradiated DNA from DiFi by pulsed-field gel electrophoresis and Southern blot hybridization using EGFR probe, indicated that the

amplified EGFR in DiFi exists in extrachromosomal, covalently-closed circular episomes, probably equivalent to dmin. Two major and one minor species were observed having estimated sizes of 650 kb, 1300 kb, and 2000 kb. The DiFi cell line appears to represent a unique case of extrachromosomal EGFR gene amplification in human cells. DiFi represents the first example of a stably maintained episome in which EGFR is amplified.

EXAMPLE 2

CONSTRUCTING A MAMMALIAN EPISOMAL EXPRESSION OR CLONING VECTOR

The identification, characterization and isolation of DNA regulatory regions within the episomes that function a) as origins of autonomous DNA replication, and b) function as stabilizing regions for extrachromosomal maintenance will permit the construction of cloning and expression vectors that replicate and function as extrachromosomal vectors. The following is meant to serve as one example of identifying and isolating such regulatory factors from the episomal unit maintained in human tumor cell. In some instances, reference is made to working with the episomal unit from DiFi cells; DiFi is used here only as an example.

I. IDENTIFICATION AND ISOLATION OF REGULATORY ELEMENTS IN STABLE EPISOMAL UNITS ESTABLISHED IN HUMAN TUMOR CELL LINES

A. Episome Isolation

In order to identify and isolate replication regulatory elements from an episome, the episome itself must first be isolated.

The ideal starting point is a preparation that is highly enriched for the episomes of interest. A highly enriched source of EGFR-containing episomes is the human

DiFi cell line. DNA will be isolated from this enriched preparation and most of the DiFi genomic DNA can be eliminated from this preparation by employing an alkaline lysis modification (Griffin, et al., J. Virol., 40:11-19, 1981). An essentially pure preparation of DiFi episomes can then be obtained by preparative electrophoresis on agarose gels that permits the mobilization of covalent circular DNA molecules (Carroll et al., Mol. Cell. Biol., 7:1740-1740 (1987)). These molecules can then be recovered from the gels by procedures that dissolve or digest (agarose) the agarose and permit the episomal DNA to be purified directly from the digest.

B. Determine a Restriction Map of the Episomal Genome.

A restriction enzyme analysis will be performed after the episome is isolated. For example, most of the DiFi episome can be separated into two pieces by exploiting the limited number of sites susceptible to restriction enzymes MluI (2 sites) and NotI (2 sites). MluI cuts at two closely spaced sites whereas NotI cuts at two widely distant sites. Table 1 presents macrorestriction fragment sizes of DiFi episomes digested with MluI and NotI restriction enzyme.

TABLE 1

MACRORESTRICTION FRAGMENT SIZES OF EPISOMES DIGESTED WITH MluI AND NotI RESTRICTION ENZYME

Restriction Enzyme	Fragment Sizes
MulI	~ 50 kb, ~ 600 kb
NotI	~ 270* kb, ~ 380** kb
MluI + NotI	~50 kb, ~220 kb, ~380 kb

* The 3′ end of this fragment contains the 5′ untranslated region, exon I, and the 5′ end of intron I of the EGFR gene.

** The 5′ end of this fragment contains the remainder of the EGFR gene from intron I through the 3′ terminus of the gene.

Digestion of total DiFi DNA with MluI and electrophoresis on agarose gels using a pulsed field gel electrophoresis format (Chu et al., Science 232:65-68, 1986) permits isolation of the region in the gel containing DNA fragments of ~600 kb. Digesting the agarose plugs with NotI further reduces the size distribution pertaining to genomic DNA and also cleaves the DiFi episome into its expected fragments. This protocol yields identifiable and highly enriched DiFi episomal fragments on a background of digested genomic DNA. The individual episomal NotI fragments (~220 and ~380 kb) are concentrated by electrophoresis in a second dimension, and then recovered from the gel by procedures that dissolve or digest agarose, thereby allowing purification of the desired DNA fragments for cloning.

C. Construction of DiFi Episome Recombinant DNA Libraries

1. **Lambda Libraries**

Lambda libraries were constructed that represented 2 to 10 kb portions of the DiFi episome by utilizing partially restriction enzyme digested episomes or NotI fragments and the Lambda-Zap phagemid vector (Short, Fernandez, Sorge, and Huse, Nuc. Acids Res. 16:7583-7600, 1988).

2. **Cosmid Libraries**

Cosmid libraries are constructed with BamHI partial digests of isolated episomes or NotI DiFi episomal fragments by utilizing the sCos1 vector (Evans, et al, Gene, 79:9-20, 1989). These cosmid libraries represent portions of the DiFi episome in approximately 40 kb blocks.

3. **P1 Libraries**

Recombinant DNA libraries containing portions of the DiFi episome are constructed by utilizing the P1 bacteriophage based cloning vector (Sternberg, Proc. Nat. Acad. Sci. USA, 87:103-107, 1990). This P1 library contains DiFi episomal portions representing two size ranges: less than 30 kb and approximately 85 - 110 kb.

4. **Plasmid libraries**

Recombinant DNA libraries containing portions of the DiFi episome are constructed utilizing an E. coli F sex factor based cloning vector (Leonardo and Sedivy, Biotechnology, 8:841, 1990). This F plasmid library contains DiFi episomal portions up to at least 150 kb. It should be understood that other plasmid libraries can be constructed using one of several available plasmid vectors (i.e. pKS, pT7\T3a-18, etc.). These vectors are known to those skilled in this art.

D. Identification of Functional Regions Within Episomes Regulating DNA Replication

In order to identify distinct episomal regions for replication, various portions of recombinant DiFi episomal DNA libraries (from the above section) are first introduced into appropriate mammalian host cells (Krysan, et. al., Mol. Cell. Biol., 9(3):1026, 1989). Autonomously replicating segments from the DiFi episome are first identified and the isolated segment is incorporated into a cloning or expression vector. Any transfection method may be employed for introducing portions of the recombinant library into mammalian host cells (i.e. calcium phosphate transfection (Chen and Okayama, Molec. Cell. Biol. 7:2745-2752 (1987)); electroporation (Chu et al., Nucl Acids Res. 15:1311-1326 (1987)).

For example, pools of approximately 10 different plasmid vector clones from the DiFi Cos1 library are introduced into for example, HSF56 human primary fibroblast cells via calcium phosphate transfection or electroporation. Each Cos1 vector clone contains a selectable marker that confers drug resistance to G418, for example. Retention and replication of transfected clones are identified by growing the transfected population of HSF56 cells in the presence of G418, a compound which specifically selects for cells that are neomycin resistant. The cells are placed under G418 selection 2 days after transfection, and G418 resistant populations are grown for at least two months by maintaining the resistant clones appropriate subculturing techniques known to those skilled in the art of tissue culture.

Neomycin resistant clones that persist for several cell divisions therefore contain a DiFi Cos1 vector clone that is replicating. A persistent neomycin resistant cell clone is recovered and low molecular weight DNA (less than 120 kb) is isolated by the HIRT extraction method (Hirt, J. Mol. Biol., 26:265-369, 1967). The DNA isolated from this neomycin resistant cell clone will be subcloned into plasmid vectors that accommodate smaller inserts, such as the pKS vector or the pT7/T3a-18 vector, which, preferably, will also contain a selectable marker, such as a gene encoding beta lactamase, which confers resistance to ampicillin.

The result of this will be another plasmid library which includes specific regions, one or more of which contain an origin for DNA replication. The clones from this new library will next be introduced into bacteria and bacterial colonies resistant to, for example, ampicillin, will be isolated. In a preferred embodiment, the host is

an E. coli cell of a type which is compatible with the vector type.

To determine if the DNA in the bacterial colonies contain an origin of DNA replication, the DNA from the ampicillin resistant bacterial colonies will be transfected into a mammalian cell line. The DNA (isolated with the HIRT extraction method) from the transfected mammalian cells will be analyzed by the Dpn I digestion (Krysan et al., Molec. Cell. Biol. 9:1026-1033, 1989 which is incorporated herein by reference). DNA exhibiting the bacterial methylation pattern is cleavable by Dpn I restriction enzyme while DNA with mammalian methylation pattern is not. Thus, DNA that is not digested by Dpn I has replicated in the mammalian cell. The origins for DNA replication will then be identified within the inserts in autonomously replicating clones. The origin can then be removed from the vector, and inserted into the recombinant cloning vector. Vectors that include regions from the DiFi episome are designated pDFE ori$^+$ and will serve as the recipients for inclusion of other regions of the DiFi episome conferring episome maintenance.

E. Identification of Functional Regions Within Episomes Regulating Extrachromosomal Maintenance

Identifying those individual clones that contain a region conferring extrachromosomal stability is determined by long term culturing (longer than two months) in the presence of a selection drug. The clones that survive the continuous exposure to the selection drug must contain a region that confers extrachromosomal stability.

Briefly, clones that persist during several cell division cycles will also be evaluated to identify regions within episomal DNA that confer stability for maintenance of extrachromosomal molecules. The procedure by which

isolation of this region is essentially the same one as described for identifying the replication region, except that vectors containing DiFi episomal origins of replication will be used to clone other restriction fragments from the DiFi episome. Once the first round of drug resistant cell colonies are identified, the episomal DNA may be isolated and introduced into bacteria and bacterial colonies resistant to, for example, ampicillin, will be isolated.

To determine if the DNA in the bacterial colonies contain a region conferring extrachromosomal stability, the DNA from the ampicillin resistant bacterial colonies will be transfected into a mammalian cell line. The DNA (isolated with the HIRT extraction method) from the transfected mammalian cells will be analyzed for fragment size and, depending on that size, another cycle may be initiated to further reduce the size of the piece of DNA that confers the extrachromosomal stability.

In addition to evidence for extrachromosomal stability that is provided by the vector's provision of drug resistance, the intranuclear localization of vector episomes will be evaluated. Vector-containing cells are treated with the non-ionic detergent Triton X-100 and 2M NaCl. This treatment produces salt extracted residual nuclei, called nucleoids, which can be centrifuged into a pellet at low speeds. Vectors associated with the nuclear matrix will pellet with the nucleoids; if they do not pellet with the nucleoids they will remain in the extracts' supernate.

Both the region for the origin of DNA replication and for extrachromosomal maintenance will be defined as the core structures of both the cloning and expression vector and will be designated PDFE ori$^+$ mat$^+$.

F. Construction of Optimal Eukaryotic Cloning Vector to Accommodate 40 kb - 400 kb Pieces of DNA.

Once the core structure is determined, construction of the optimal eukaryotic cloning or expression vector will be completed. This is accomplished by adding the following three features to the core structure (these will be discussed below):

a. a DNA or genomic DNA region encoding at least one selectable marker;

b. a DNA or genomic DNA region encoding a multiple cloning site; and

c. a DNA or genomic DNA region encoding bacterial components necessary for propagation of the vector in bacteria.

Selectable markers, for mammalian cells, confer resistance to a specific selection agent once DNA conferring the resistance is transfected into individual cells possessing a genetic inheritance pattern appropriate for the selectable marker being used in the vector. There are a variety of different dominant and recessive selection agents known to those skilled in the art. Any one of the following genes and agents should be effective in terms of employing a selection system:

- G418 resistance is selected by exposure to medium containing 100 to 800 ug/ml G418. G418 selects for cells deficient in the enzyme aminoglycoside phosphotransferase and are referred to as neomycin resistant cells. (Southern and Berg, J. Molec. Appl. Gen.,

1:327-341, 1982; Colbere-Garapin et al., J. Molec. Biol., 150:1, 1981).

- HAT resistance for forward selection (converting a thymidine kinase minus cell to a thymidine kinase positive cell) is selected with complete medium supplemented with 100 uM hypoxanthine, 0.4 uM aminopterin, 16 uM thymidine and 3 uM glycine. HAT medium selects for variants defective in either hypoxanthine-guanine phosphoribosyl-transferase or thymidine kinase (Littlefield, Proc. Natl. Acad. Sci. USA, 50:568, 1963; Littlefield, Science, 145:709-710, 1964).

- Hygromycin B resistance is selected by exposure to complete medium supplemented with 10 - 400 ug/ml hygromycin B. Hygromycin B selects for variants defective in the enzyme hygromycin-B-phosphotransferase (Gritz and Davies, Gene, 25:179-188, 1983; Santerre, et al., Gene, 30:147, 1984; Palmer, et.al., Proc. Natl. Acad. Sci. USA, 84:1055-1059, 1987).

- Adenine phosphoribosyltransferase (APRT) positive variants are selected by exposure to medium supplemented with 25 uM alanosine, 50 uM azaserine and 100 uM adenine (Lowy, et. al., Cell, 22:817, 1980; Adair, et. al., Proc. Natl. Acad. Sci. USA, 86:4574-4578, 1989).

- Xanthine - Guanine Phosphoribosyltransferase (XGPRT) positive variants are selected with complete medium supplemented with dialyzed fetal calf serum, 250 ug/ml xanthine, 15 ug/ml hypoxanthine, 10 ug/ml

thymidine, 2 ug/ml aminopterin, 25 ug/ml mycophenolic acid, and 150 ug/ml L-glutamine (Mulligan and Berg, Proc. Natl. Acad. Sci. USA, 78:2072-2076, 1981).

- Methotrexate resistance is selected by exposure to complete medium supplemented with 0.01 uM - 300 uM methotrexate and dialyzed fetal calf serum. Methotrexate selects for cells expressing high levels of dihydrofolate reductase (O'Hare, et al., Proc. Natl. Acad. Sci. USA, 78:1527, 1981; Simonsen and Levinson, Pro. Natl. Acad. Sci. USA, 80:2495-2499, 1983).

- Deoxycoformycin resistant cells are selected by exposure to complete medium supplemented with 10 ug/ml thymidine, 15 ug/ml hypoxanthine, 4 uM 9-B-D- xylofuranosyl adenine (XylA), and 0.01 - 0.03 uM 2'-deoxycoformycin (dCF). This selection selects for mutants expressing adenosine deaminase (ADA; Kaufman, et. al., Proc. Natl. Acad. Sci. USA, 83:3136-3140, 1986).

For added ease in handling and manipulating, this optimum eukaryotic cloning vector could include a DNA region comprising a multiple cloning cassette sequence containing infrequent cutting by restriction enzymes to facilitate the insertion of a desired gene. Multiple cloning cassette sequence cartridges are commercially available from several different companies (*Stratagene*, Promega, New England Biolabs etc). A typical cassette sequence cartridge would include restriction sites for 8 - 11 different enzymes (i.e. Eco R1, Sac1, Sma 1, Ava 1, Bam H1, Xba 1, Hinc II, Acc 1, Sal 1, Pst 1, Hind III, etc.).

The availability of these cassette cartridges are known to those skilled in the art.

The bacterial plasmid sequences may be derived from any one of the many different vectors that are commercially available and known to those skilled in the art of recombinant DNA technology. For the purpose of this invention, pUC, pKS, pBR322 and pT7/T3a18 are used as a matter of preference, however, other vectors would be equally effective. For example, if pBR322 sequences are introduced into the cloning or expression vector, the resulting recombinant can then be shuttled back and forth between E. coli and mammalian cells.

The construction of an optimal eukaryotic expression vector that can accommodate 40 kb - 400 kb pieces of DNA will also contain, in addition to the elements described for the cloning vector, a DNA region containing a promoter, a polyadenylation and splice site necessary for the expression of the desired gene.

There are at least two approaches for constructing an optimal eukaryotic cloning vector that can accommodate 40 - 400 kb pieces of DNA.

1. The first and more simpler approach is to begin with a readily available cloning plasmid vector capable of propagation in bacteria. There are many different vectors known to those skilled in the art that would work efficienty. Several different components and features can easily be ligated into this bacterial plasmid vector. These added features are discussed below. Once completed, the vector will not only have the core structure (to confer the ability to replicate DNA and to be maintained

extrachromosomally) but will also have the added features to optimize the vector for propagation in bacteria and for identification of its presence after transfection into a mammalian cell recipient.

2. The second approach involves custom designing andcreating the optimum cloning vector by ligating all the desired features and components (including the core structure) together to generate the vector of choice.

G. Construction of a Mammalian Artificial Chromosome

The episomally maintained and replicated vector pDFE ori$^+$ mat$^+$ is introduced into cells and persist as covalent circular extrachromosomal molecules. In this form the episomes accumulate to produce multiple copies in each cell and accordingly, also overproduce mRNA and its protein product. While this is desirable for producing amplified genes and gene products, the introduction of cloned genes into cells for use in gene therapy requires the control of gene copy number and attendant gene expression. Such control is introduced into the DiFi episome vector by introducing DNA sequences that stabilize artificial chromosomes containing linear double stranded DNA (DNA encoding a telomere). Such sequences occur at the termini of natural chromosomes; in human chromosomes 5′-AGGGTT-3′ is tandemly repeated to the extent of 10 of 15 kb at every telomere (Blackburn, Science, 249:489, 1990). This tandemly repeated sequence is ligated to each end of a linearized cloning and expression vector to stabilize the termini. The addition of telomere sequences specific for other species provides for the stabilization of artificial chromosomes when introduced into those species.

Centromere sequences are known to identify regions within chromosomes where kinetichores are organized and mitotic spindles are attached to the chromosomes, thus ensuring for the segregation of chromosomes during mitosis. DNA sequences that serve as centromeres are introduced into an internal region of the linearized cloning and expression vector which contain telomeres resulting in an artificial chromosome. This synthetic chromosome contains required regulatory and stabilizing DNA sequences that normally occur in natural chromosomes.

Specific genetic function is conferred on this synthetic chromosome by ligating a gene of interest into its multiple cloning site. For example, the gene or cDNA derivative of the gene that is defective in Duchenne's muscular dystrophy or myotonic dystrophy, or one of a number of other diseases associated with muscle dysfunction may be cloned into the artificial chromosome. The artificial chromosome is then introduced into cells or tissues or animals by methods appropriate for the target. The transfected chromosome is established as an integral component of the recipient cells where it is stably maintained and expressed. Recipient cells, tissues or animals that were initially dysfunctional because of a genetic defect they possessed are cured and become normal because of the expression and synthesis of the normal gene product introduced in the artificial chromosome.

H. Evaluation of Different Strategies for Transfecting Cloning or Expression Vectors Into Mammalian Cells

Once the optimal cloning and expression vector is constructed, several different strategies for transfecting the vectors will be studied. Examples of potential methods includes: (1) encapsulation of insert-containing vectors in liposomes of appropriate composition to enhance

entry into target cells, and (2) electroporation of vector into mitotic cell recipients to enhance its inclusion within the nucleus as cells progress into G1 phase of the cell cycle, and (3) injection of DNA-encoated particles into cells by employing a Biolistic Particle Delivery System (DuPont). This procedure essentially shoots DNA-coated bullets into cells or tissues.

I. Biosynthetic Production of Proteins in Cells Transfected With Cloning and Expression Vectors Containing Isolated Genes or Functional Derivatives

Medically important proteins are produced in mammalian cells that have been transfected with the vector containing the gene encoding the protein. Since the gene-containing vector accumulates in the transfected cells, the amount of protein produced increases as more vector copies accumulate. The following example illustrates an efficient system for protein production. To produce the product of the gene that is deficient in patients with myotonic dystrophy, the vector containing the normal gene is electroporated into a normal primary human fibroblast cell line HSF56, adapted for growth in suspension culture in serum free medium. The accumulation of the cloning vector in each cell is accelerated by growing the cells in the drug appropriate for the drug resistance gene contained in the vector. As the gene copy number accumulates the amount of protein increases to be recovered from the culture medium or from the cells after maximal growth is achieved. The medical condition of patients with myotonic dystrophy may be improved by treatment with the protein that is provided by this cloningexpression system. Modification of the vector to include other genes and selection of target cells and appropriate culture conditions provides endless possible

systems for the production and isolation of mammalian proteins.

The foregoing description has been directed to particular embodiments of the invention in accordance with the requirements of the Patent Statutes for the purposes of illustration and explanation. It will be apparent, however, to those skilled in this art, that many modifications and changes in the apparatus and procedure set forth will be possible without departing from the scope and spirit of the invention. It is intended that the following claims be interpreted to embrace all such modifications and changes.

CLAIMS

1. A composition of matter comprising a substantially purified non-viral episome of human origin capable of stable extrachromosomal maintenance and of autonomous replication in a compatible mammalian cell line.

2. A substantially purified episomal DNA segment containing an origin of replication.

3. A substantially purified episomal DNA segment containing a DNA sequence which confers upon a vector including said segment the ability to be stably maintained extrachromosomally in a cell transfected with said vector.

4. A substantially purified episomal DNA segment containing an origin of replication and a DNA sequence which confers upon a vector including said segment the ability to be stably maintained extrachromosomally in a cell transfected with said vector.

5. The substantially purified episomal DNA segment of claim 2, 3, or 4 wherein the episomal DNA segment is from an episome isolated from DiFi colorectal cell line.

6. A cloning vector comprising the following components operatively spaced with respect to a desired gene:

a) a DNA segment derived from a non-viral episome, said segment containing an origin for DNA replication;

b) a DNA segment derived from a non-viral episome, said segment containing a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector;

c) a DNA segment containing a multiple cloning site;

d) a DNA selectable marker segment conferring upon a cell transfected with said vector the ability to survive in the presence of a selected compound or selected group of compounds; and

e) a DNA segment encoding bacterial components necessary for propagation of said vector in bacteria.

7. The cloning vector of claim 6 wherein said compound is selected from the group consisting of G418 and hygromycin B.

8. The cloning vector of claim 6 further including a DNA sequence encoding a desired protein.

9. The cloning vector of claim 6 wherein the segment containing the origin for DNA replication is from an episome isolated from DiFi colorectal cell line.

10. The cloning vector of claim 6 wherein the segment containing a DNA sequence which confers upon said

vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector is from an episome isolated from DiFi colorectal cell line.

11. The cloning vector of claim 6 wherein the chromosomal DNA of said transfected cell contains a mutation in an enzyme, said mutation rendering said selected compound or selected group of compounds toxic to said cell when said selectable marker segment is not present in said cell and wherein said selectable marker segment contains DNA encoding an enzyme capable of functionally replacing said mutated enzyme so as to render said transfected cell resistant to said selected compound or selected group of compounds.

12. The cloning vector of claim 11 wherein said enzyme is selected from the group consisting of: thymidine kinase, xanthine-guanine phosphoribosyl transferase, adenine phosphoribosyltransferase, adenosine deaminase and dihydrofolate reductase.

13. A cloning vector comprising the following components operatively spaced with respect to a desired gene:

a) a DNA segment derived from a non-viral episome, said segment containing an origin for DNA replication and a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector;

b) a DNA segment containing a multiple cloning site;

c) a DNA segment conferring upon a cell transfected with said vector the ability to survive in the presence of a selected compound or selected group of compounds; and

d) a DNA segment encoding bacterial components necessary for propagation of said vector in bacteria.

14. The cloning vector of claim 13 wherein said compound is selected from the group consisting of G418 and hygromycin B.

15. The cloning vector of claim 13 further including a DNA sequence encoding a desired protein.

16. The cloning vector of claim 13 wherein the DNA segment containing an origin for DNA replication and a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector is from an episome isolated from DiFi colorectal cell line.

17. The cloning vector of claim 13 wherein the chromosomal DNA of said transfected cell contains a mutation in an enzyme, said mutation rendering said selected compound or selected group of compounds toxic to said cell when said selectable marker segment is not present in said cell and wherein said selectable marker

segment contains DNA encoding an enzyme capable of functionally replacing said mutated enzyme so as to render said transfected cell resistant to said selected compound or selected group of compounds.

18. The cloning vector of claim 17 wherein said enzyme is selected from the group consisting of: thymidine kinase, xanthine-guanine phosphoribosyl transferase, adenine phosphoribosyltransferase, adenosine deaminase and dihydrofolate reductase.

19. An expression vector comprising the following components operatively spaced with respect to a desired gene:

a) a DNA segment derived from a non-viral episome, said segment containing an origin for DNA replication;

b) a DNA segment derived from a non-viral episome, said segment containing a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector;

c) a DNA segment containing a multiple cloning site;

d) a DNA segment conferring upon a cell transfected with said vector the ability to survive in the presence of a selected compound or selected group of compounds;

e) a DNA segment encoding bacterial components necessary for propagation of said vector in bacteria; and

f) a promoter, a polyadenylation site, and a splice site in spacial relation to allow the efficient expression of a structural gene upon insertion of said gene into said splice site.

20. The expression vector of claim 19 wherein said compound is selected from the group consisting of G418 and hygromycin B.

21. The expression vector of claim 19 further including a DNA sequence encoding a desired gene.

22. The expression vector of claim 19 wherein the bacterial components necessary for propagation of said vector in bacteria are derived from pBR322, pUC, pT7/T3a-18 or pKS.

23. The expression vector of claim 19 wherein the segment containing the origin for DNA replication is from an episome isolated from DiFi colorectal cell line.

24. The expression vector of claim 19 wherein the segment containing a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector is from an episome isolated from DiFi colorectal cell line.

25. The expression vector of claim 19 wherein the chromosomal DNA of said transfected cell contains a mutation in an enzyme, said mutation rendering said selected compound or selected group of compounds toxic to said cell when said selectable marker segment is not present in said cell and wherein said selectable marker segment contains DNA encoding an enzyme capable of functionally replacing said mutated enzyme so as to render said transfected cell resistant to said selected compound or selected group of compounds.

26. The expression vector of claim 25 wherein said enzyme is selected from the group consisting of: thymidine kinase, xanthine-guanine phosphoribosyl transferase, adenine phosphoribosyltransferase, adenosine deaminase and dihydrofolate reductase.

27. An expression vector comprising the following components operatively spaced with respect to a desired gene:

a) a DNA segment derived from a non-viral episome, said segment containing an origin for DNA replication and a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector;

b) a DNA segment containing a multiple cloning site;

c) a DNA segment conferring upon a cell transfected with said vector the ability to survive in the

presence of a selected compound or selected group of compounds;

d) a DNA segment encoding bacterial components necessary for propagation of said vector in bacteria; and

e) a promoter, a polyadenylation site, and a splice site in spacial relation to allow the efficient expression of a structural gene upon insertion of said gene into said splice site.

28. The expression vector of claim 27 wherein said compound is selected from the group consisting of G418 and hygromycin B.

29. The expression vector of claim 27 further including a DNA sequence encoding a desired protein.

30. The expression vector of claim 27 wherein the bacterial components necessary for propagation of said vector in bacteria are derived from pBR322, pUC, pT7/T3a-18 or pKS.

31. The expression vector of claim 27 wherein the DNA segment containing an origin for DNA replication and a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector is from an episome isolated from DiFi colorectal cell line.

32. The expression vector of claim 27 wherein the chromosomal DNA of said transfected cell contains a mutation in an enzyme, said mutation rendering said selected compound or selected group of compounds toxic to said cell when said selectable marker segment is not present in said cell and wherein said selectable marker segment contains DNA encoding an enzyme capable of functionally replacing said mutated enzyme so as to render said transfected cell resistant to said selected compound or selected group of compounds.

33. The cloning vector of claim 32 wherein said enzyme is selected from the group consisting of: thymidine kinase, xanthine-guanine phosphoribosyl transferase, adenine phosphoribosyltransferase, adenosine deaminase and dihydrofolate reductase.

34. The expression vector of claim 19 or 27 wherein the promoter is selected from the group consisting of: cytomegalovirus promoter, SV-40 promoter, Rous sarcoma virus promoter, thymidine kinase promoter, beta-actin promoter, metallothionein promoter, and epidermal growth factor receptor gene promoter isolated from a DiFi episome.

35. An artificial chromosome comprising:

a DNA segment derived from a non-viral episome, said segment containing an origin for DNA replication, a DNA segment derived from a non-viral episome, said segment containing a DNA sequence which confers upon said vector the ability to be stably maintained extrachromosomally in a cell transfected with said vector, a DNA segment containing a multiple cloning site, a DNA selectable

marker segment conferring upon a cell transfected with said vector, the ability to survive in the presence of a selected compound or selected group of compounds, a DNA segment encoding bacterial components necessary for propagation of said vector in bacteria, a promoter, a polyadenylation site, a splice site, a DNA segment encoding a centromere and a DNA segment encoding a telomere.

36. The artificial chromosome of claim 35 wherein said compound is selected from the group consisting of G418 and hygromycin B.

37. The artificial chromosome of claim 35 further including a DNA sequence encoding a desired protein.

38. The artificial chromosome of claim 35 wherein the chromosomal DNA of said transfected cell contains a mutation in an enzyme, said mutation rendering said selected compound or selected group of compounds toxic to said cell when said selectable marker segment is not present in said cell and wherein said selectable marker segment contains DNA encoding an enzyme capable of functionally replacing said mutated enzyme so as to render said transfected cell resistant to said selected compound or selected group of compounds.

39. The artificial chromosome of claim 38 wherein said enzyme is selected from the group consisting of: thymidine kinase, xanthine-guanine phosphoribosyl transferase, adenine phosphoribosyltransferase, adenosine deaminase and dihydrofolate reductase.

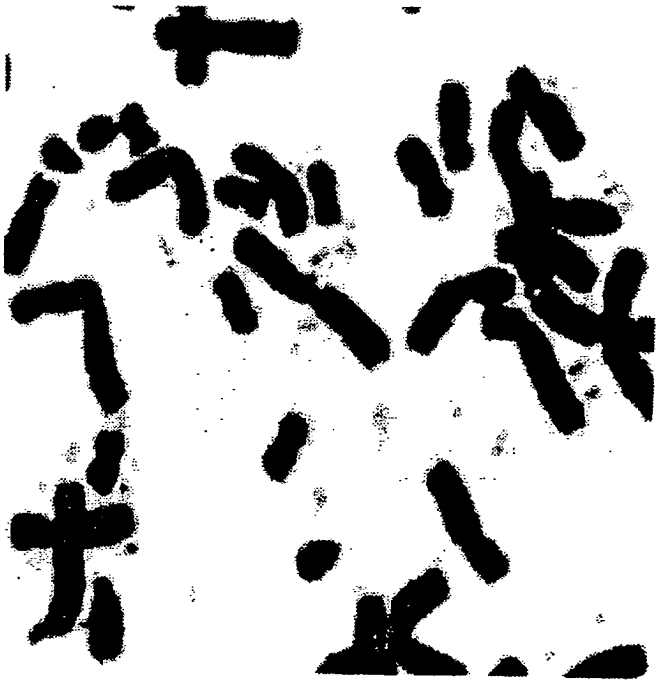


FIG. 1A

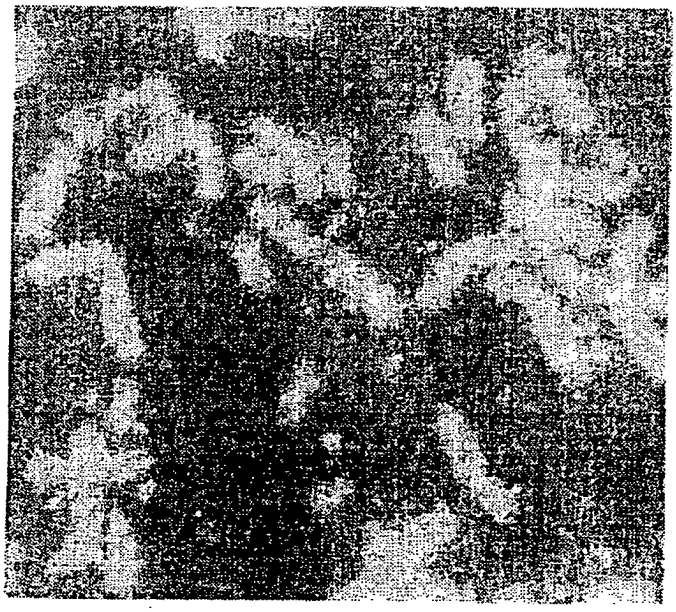


FIG. 1B

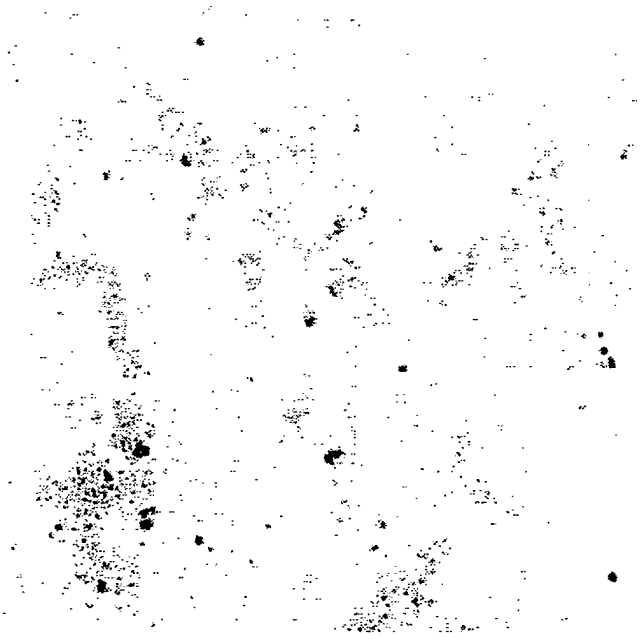
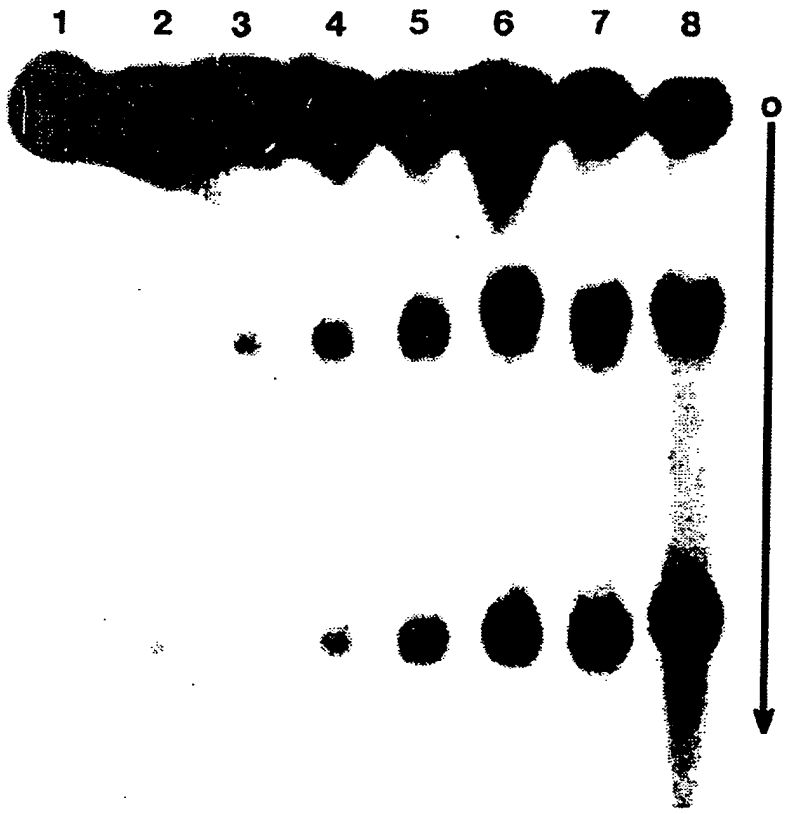


FIG. 1C



O - 320 GRAY

FIG. 2

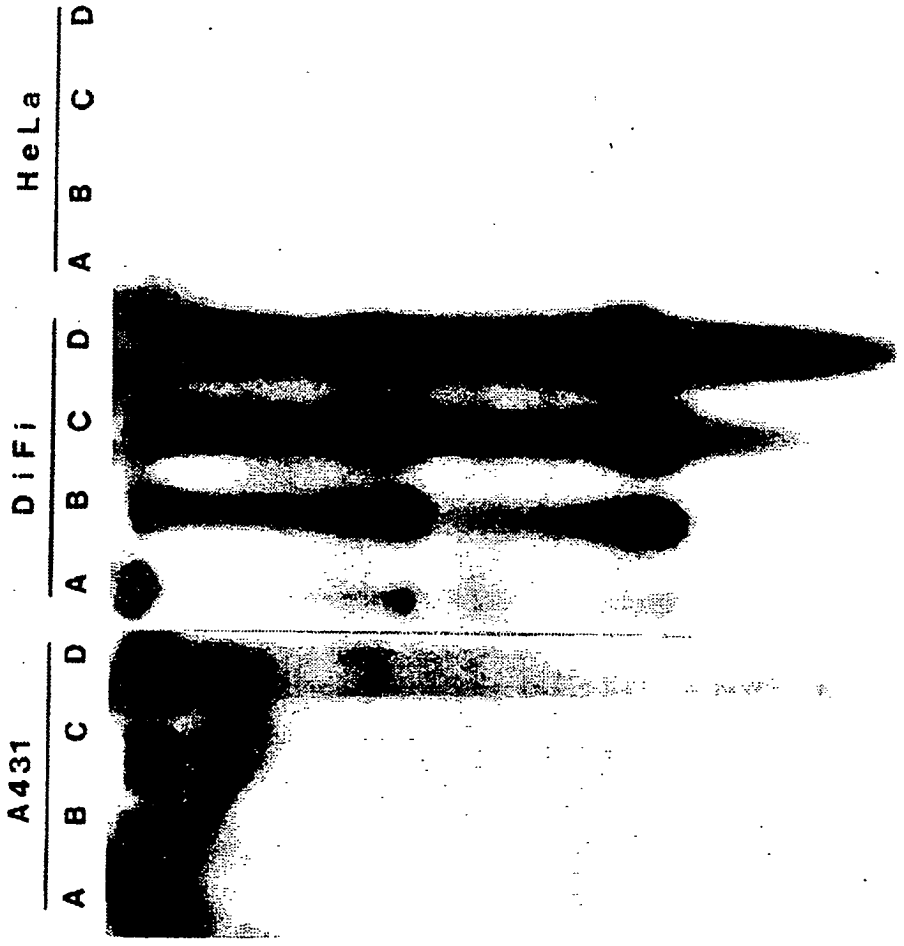


FIG. 3

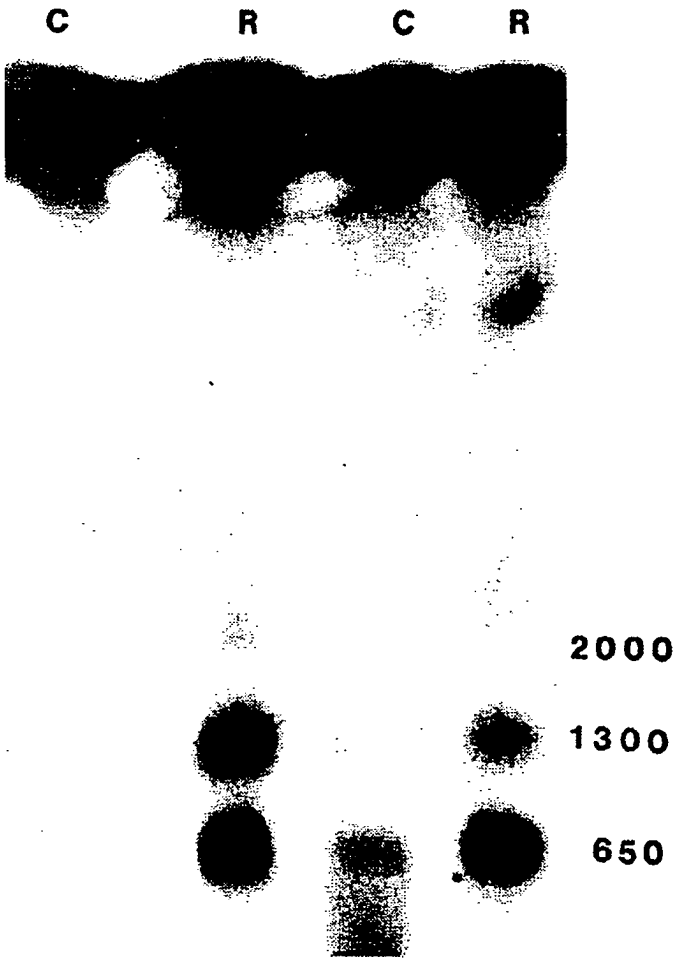


FIG. 4

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 91/07690

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/85; C12N15/87

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category *	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
X	AM. J. HUM. GENET. vol. 45, no. 4 SU, 1989, page A22; GOODE, M.E. ET AL.: 'Molecular characterization of egfr gene amplification in the colon carcinoma cell line DiFi' see abstract (0080) 7.10 ---	1-5
P,X	GENES, CHROMOSOMES & CANCER vol. 3, no. 1, 13 February 1991, pages 48 - 54; DOLF, G. ET AL.: 'Extrachromosomal amplification of the epidermal growth factor receptor gene in a human colon carcinoma cell line' see the whole document ---	1-5
X	EP,A,0 254 315 (DAIICHI SEIYAKU) 27 January 1988 see the whole document ---	1-4
	-/--	

* Special categories of cited documents:[10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
21 FEBRUARY 1992	04. 03. 92
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	CHAMBONNET F.J.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X	EP,A,0 350 052 (DAIICHI PHARMACEUTICAL LTD) 10 January 1990 see the whole document	1-4,6,8, 13,15, 19,21, 27,29,34
X	EP,A,0 306 848 (BOEHRINGER MANNHEIM GMBH) 15 March 1989 see the whole document	1-4,6,8, 11-13, 15, 17-19, 21,25-26

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 9107690 SA 53335

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information. 21/02/92

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A-0254315	27-01-88	JP-A- 63185386	30-07-88
EP-A-0350052	10-01-90	JP-A- 2131582	21-05-90
		JP-A- 2131583	21-05-90
		JP-A- 2186931	23-07-90
EP-A-0306848	15-03-89	DE-A- 3730246	01-06-89
		JP-A- 1098492	17-04-89